COVID-19 (Stay at home, don't let the virus roam)

-By Rajkumar Gosai

COVID-19 also called as Novel Corona Virus emerged in China in late-2019's and has spread all over the world in just couple of months. The main reason of the widespread of this disease is that it spreads though human contact and it can remain dormant in human body from 10 to 20 days and scientist are still struggling to develop a vaccine for this disease.

Aim:

While there is not stopping in the increasing number of COVID-19 in most countries, we will try to analyze and visualize the countries that are hit badly due to this virus. We will also visualize the number of confirmed. Deaths and recovered cases due to COVID-19 using the data that we collected through this presentation.

About Dataset:

This dataset is accumulated by John Hopkins University in the form of google sheets and we are using these google for our analysis. There are total 8 csv files present in the dataset. We will be using 5 of these files which are "covid_19_data", "time_series_covid_19_confirmed", "time_series_covid_19_deaths", "time_series_covid_19_confirmed_US" and "time_series_covid_19_deaths_US" for our analysis.

This dataset has covid-19 cases recorded across different dates starting from 22nd January to 10th April 2020.

You can find the dataset from the link: https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

Analysis and Visualizations:

As our dataset has covid-19 cases from all around the world, we will be using geo map to plot the total number of confirmed cases in the countries present in the dataset. The varying difference in the total number of confirmed cases in different countres is shown by the changing color pattern scheme.

Also, I have extracted shape file as the COVID-19 file did not had iso-166 3-digit country code required to plot geo map in my case. The shape file was merged with the "covid_19_data" file in order to plot the covid data on the geo map.

Geo map makes it very easy to find the countries that are struck badly by covid-19 virus by use of different color schemes. The top countries affected by this virus are: US, Spain, Italy, China, France, Germany, Iran, Iraq.
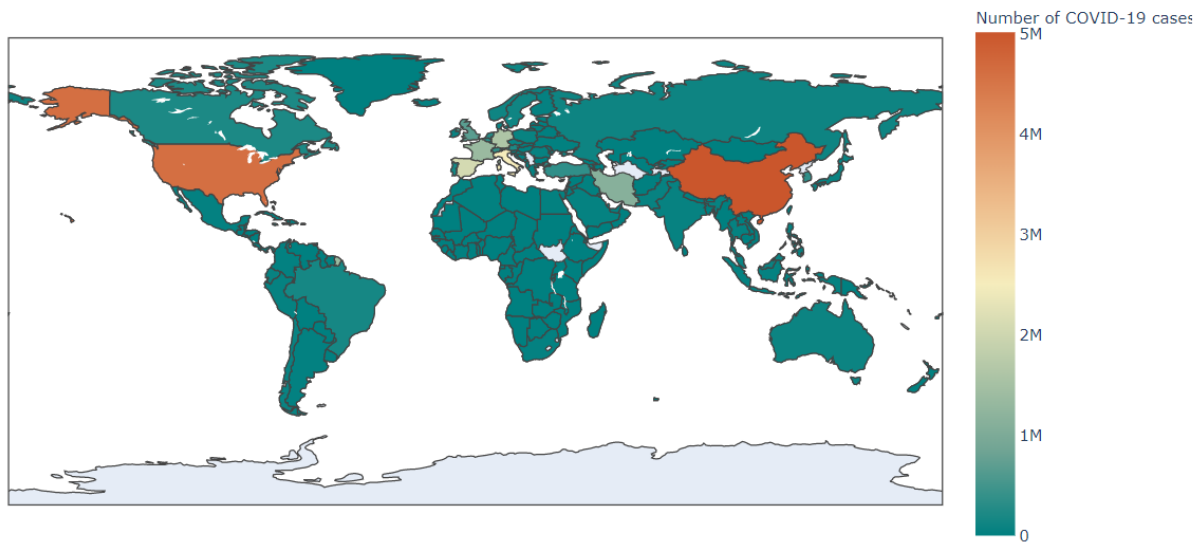
Now in order to see the total number of confirmed, deaths and recovered cases for a country using the geo-map when clicked on, a line graph was linked with the map graph which displays the total number of cases for the given country.

For example, we want to see the total number of cases in China, we will simply click on the country and it will extract the data of China and plot the cases accordingly.

Joint scatter plot showing Confirmed, Deaths and Recoverd cases of China
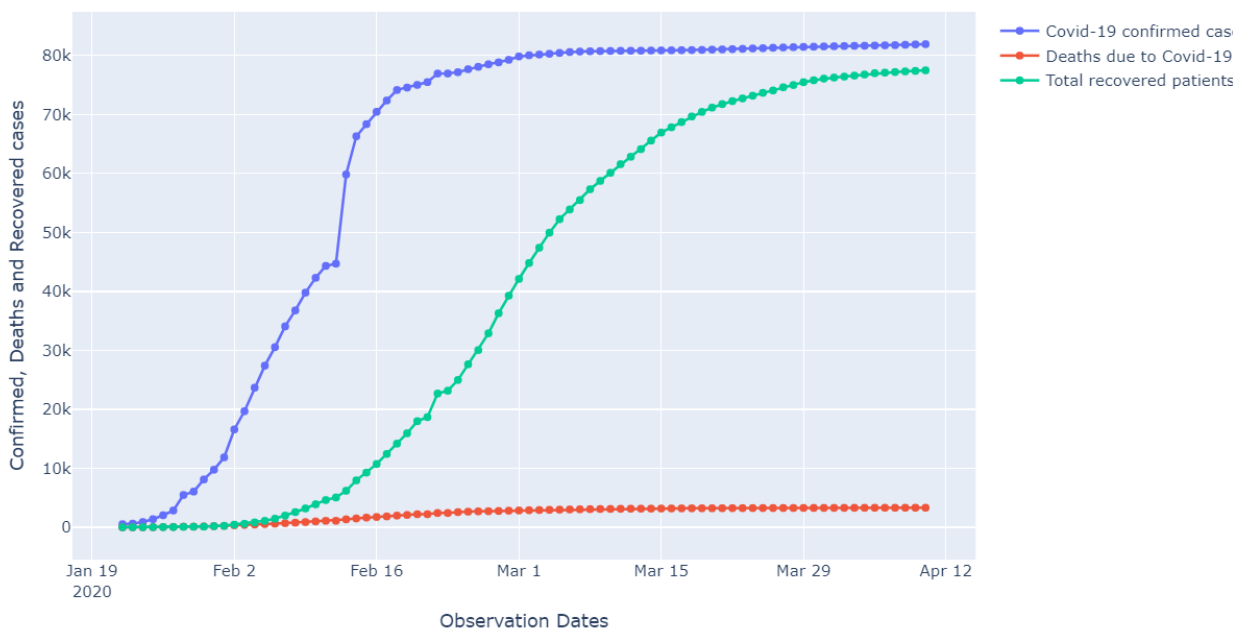
Insights: There was rapid increase in the number of COVID cases on 12th Feb. The total number of cases went from around 45k to 60k in just a single day.

From the above graph, we can see that China has pretty good recovery record (around 95%) when compared to other countries around the world like the US for example, as it is evident from the below graph

Joint scatter plot showing Confirmed, Deaths and Recoverd cases of United States of America
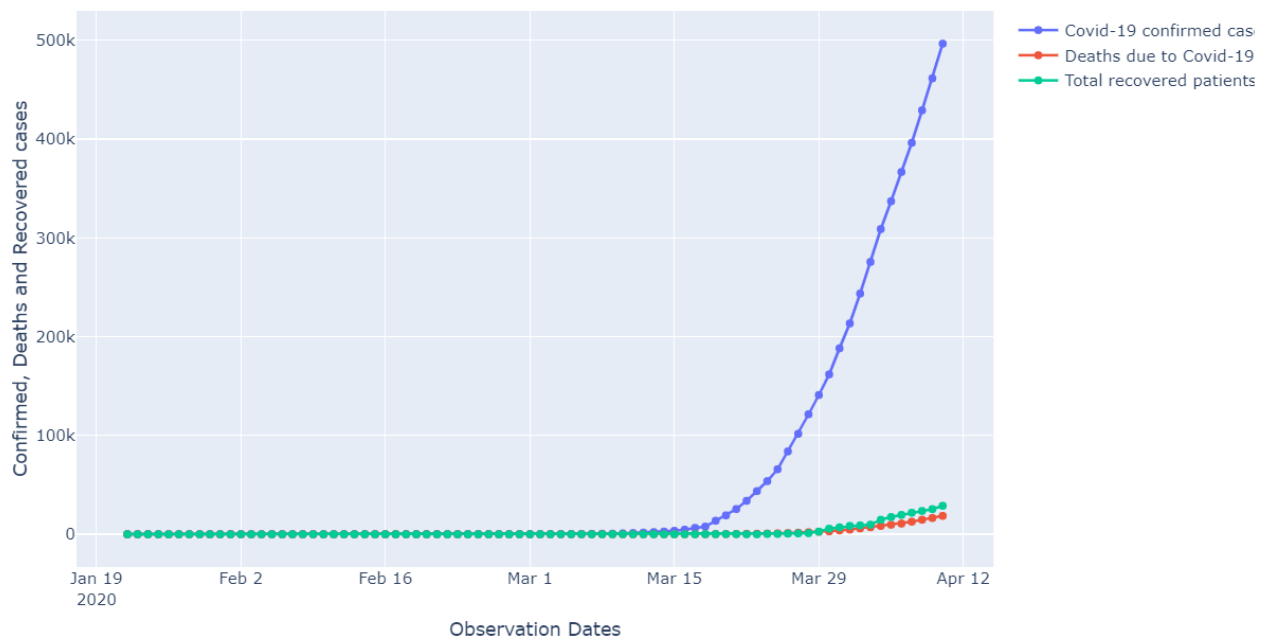

Country-wise Interactive Graph:

Below is a country-wise interactive graph visualization because there are some countries that might by bit difficult to trace in the map because of the small map size. So, below graph shows the county-wise visualization when selected.

COVID-19 cases in Spain
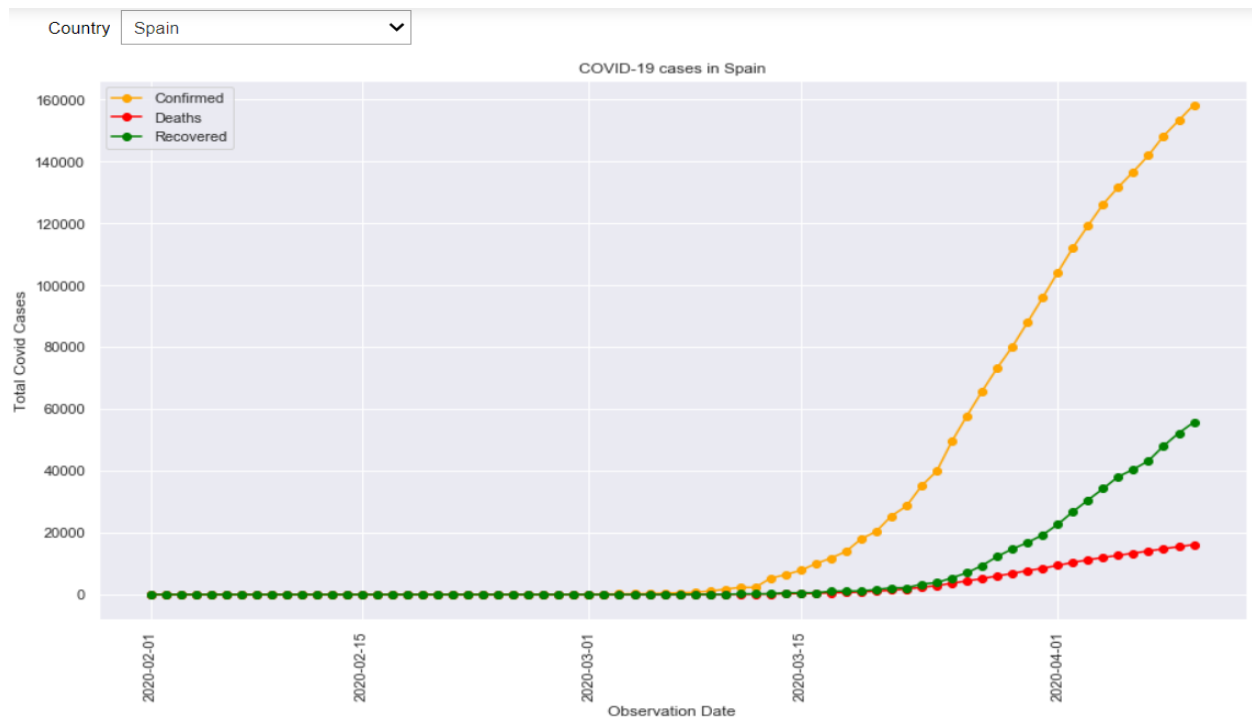
Insights: Spain is the second country having the greatest number of covid cases. The total number of COVID cases in Spain is around 160k and the recovery rate is comparatively better than that of US.

As US has most number of COVID-19 cases till date and "New York" is one of the worst hit cities, we tried to visualize the US data state wise and you can see in the graph below.

```
New York : Red
New Jersey : Green
Michigan : Blue
California : Yellow
Massachusetts : Magenta
```


Datewise confirmed COVID cases in top 10 states in United States of America
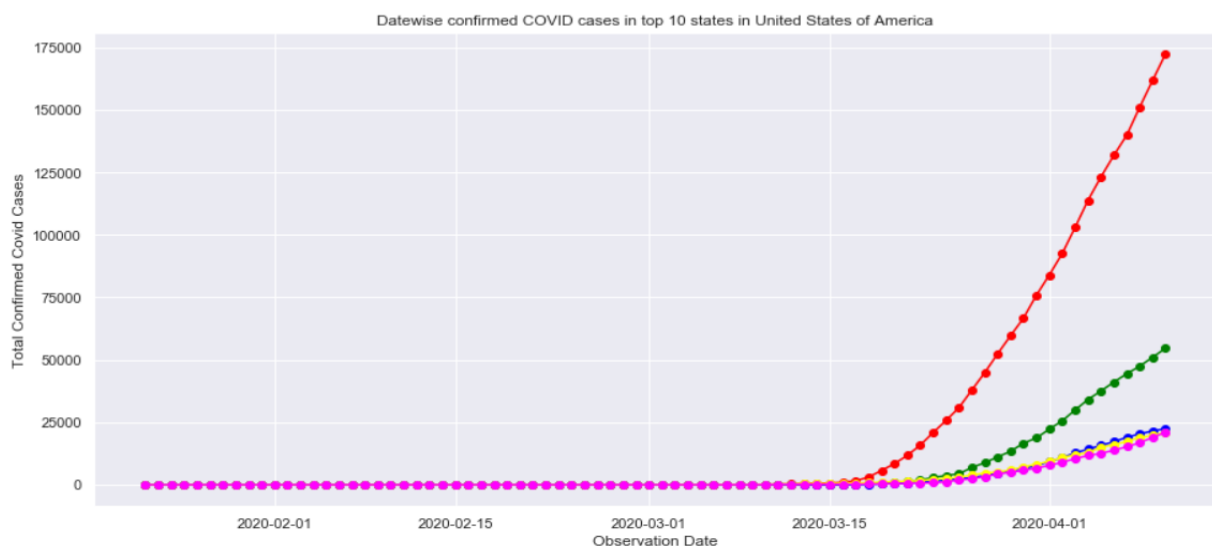
Insights:

The total number of confirmed cases in New York itself is around 175k which is more than the total number of covid cases in second worst hit country Spain (160k). Which is followed by New Jersey, Michigan etc.

The total number of death cases in New York is almost less than half than that of Spain as you can see from the below histogram. (taken from Assignment Part #1 Visualization)

Total number of deaths by Cities
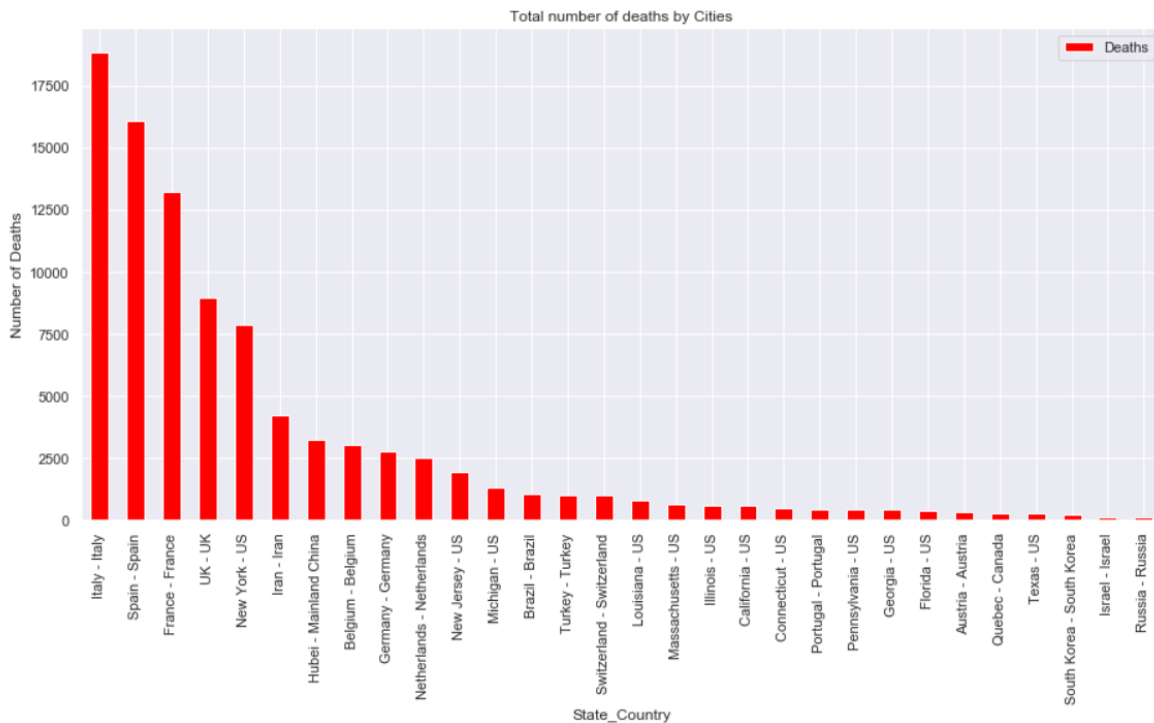
When we look at Death Percentage over total number of the covid cases, Italy has the highest number of death percentage with an over 12% death ratio followed by UK and Belgium respectively.

This high death percentage might vary in the sense that 50% of population in Italy who were infected the virus were above 60 years of age and were already suffering from some kind of terminal disease as given in the article https://www.livescience.com/why-italy-coronavirus-deaths-so-high.html

Total Deaths % by Cities
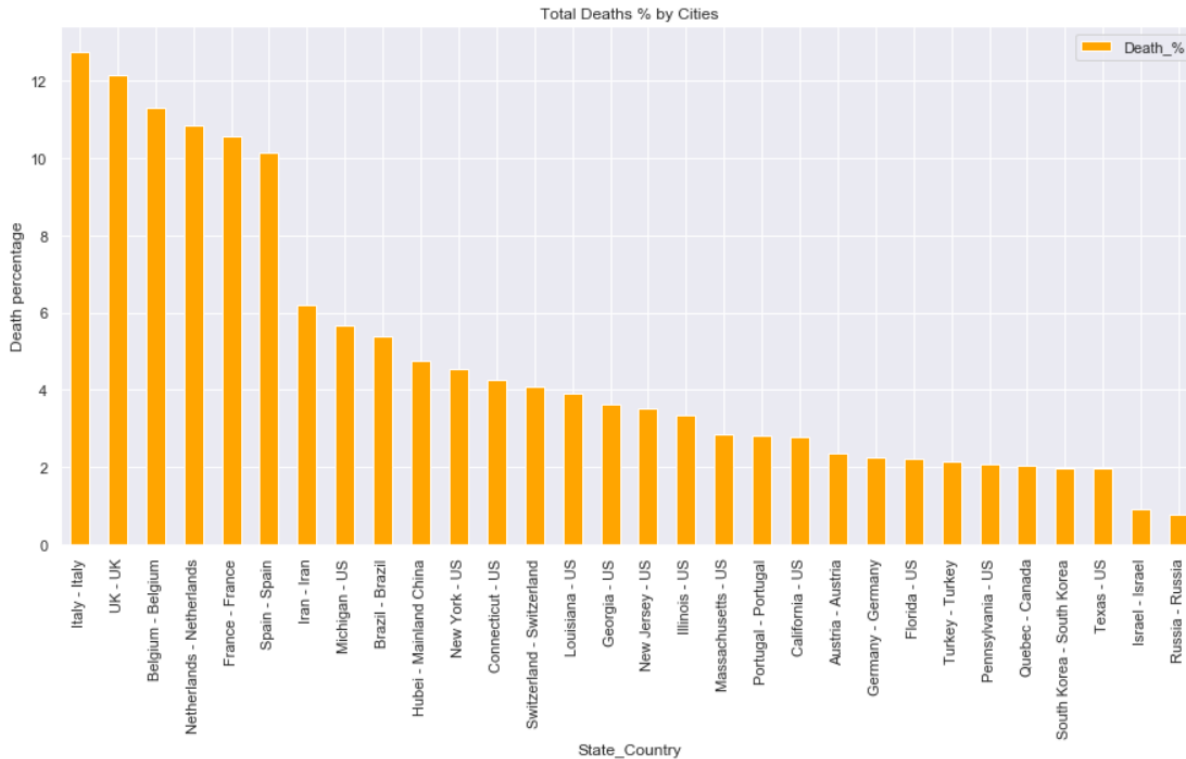
Conclusion:

1. US has the greatest number of COVID-19 cases than any other country in the world.
2. New York is the worst hit state in the US by COVID-19 where the total number of cases in a single state is higher than the total number of cases in the Spain which is the second worst hit country by the virus
3. Total number of deaths is much higher than that in New York
4. Italy has the highest number of death rate percentage with over 12%
5. There are other factors as well that lead to increasing number of deaths that remain accounted in the data collected

References:

1. https://plotly.com/python/reference/#choropleth
2. https://plotly.com/python/figure-labels/
3. https://plotly.com/python/choropleth-maps/
4. https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset
5. https://bqplot.readthedocs.io/en/latest/_generate/bqplot.scales.GeoScale.html
6. https://plotly.com/python/figurewidget/
7. https://wwwnc.cdc.gov/eid/article/26/7/20-0407_article
8. https://www.gazettetimes.com/news/opinion/mailbag/mailbag-covid-19-needs-its-own-catchphrase/article_59ee6208-61c3-5219-8154-cf2d60de581e.html